**Scientific and Practical Cyber Security Journal (SPCSJ) 6(2): 23-29 ISSN 2587- 4667 Scientific Cyber Security Association (SCSA)**

# A SURVEY ON KNOWLEDGE AND COMMONSENSE REASONING FOR NATURAL LANGUAGE PROCESSING

**Aliyu Ahmed Abubakar Kaduna State University Wuhan University**

**ABSTRACT:** People use knowledge and commonsense reasoning for daily activities and survival. However, providing machines with such humanly knowledge and commonsense reasoning experiences has remained a vague target of artificial intelligence researchers for years. This report surveys knowledge and commonsense reasoning for Natural Language Processing with the aim of providing an overview of the benchmarks, knowledge resources, state of the art and inference approach toward knowledge and commonsense reasoning for natural language processing.

**KEYWORDS –** *Commonsense Reasoning, Knowledge Resource, Natural Language Processing (NLP), Artificial Intelligence (AI)*

## 1. Introduction

Knowledge and commonsense reasoning is the cornerstone of the application of human intelligence according to (Razniewski, Tandon, & Varde, 2021). Knowledge and commonsense reasoning in artificial intelligence (AI) is a human-like ability to make assumptions regarding the kind and essence of normal situations humans encounter daily. These presumptions include decisions regarding the nature of peoples' intentions, physical objects and taxonomic properties. Knowledge and commonsense reasoning is relevant for several applications of current interest and such applications include robot and human collaboration, transparent machine-learning systems which will be able to explain their conclusions, dialogue systems, social media and story understanding software.

With the speedy improvement of Human Computer Interactions engines (such as chat, dialogue systems and QA), making use of knowledge and commonsense reasoning in natural language understanding has become a very important area in NLP, as they are necessary for conversation engines or other sorts of HCI engines to comprehend user queries, manage conversations, as well as generating responses (Zhou, Duan, Wei, Liu, & Zhang, 2018). Knowledge and commonsense reasoning have acquired repeated consideration from the natural language processing (NLP) community recently, resulting numerous exploratory research directions into automated commonsense understanding (Maarten, Vered, Antoine, Yejin, & Dan, 2020). Devlin, Chang, Lee, & Toutanova, (2019); Liu, et al. (2019), have lately, made lots of advances in large pre-trained language models where they tried pushing machines nearer to humanlike understanding capabilities, making researchers wonder if machines could directly model commonsense through symbolic integrations.

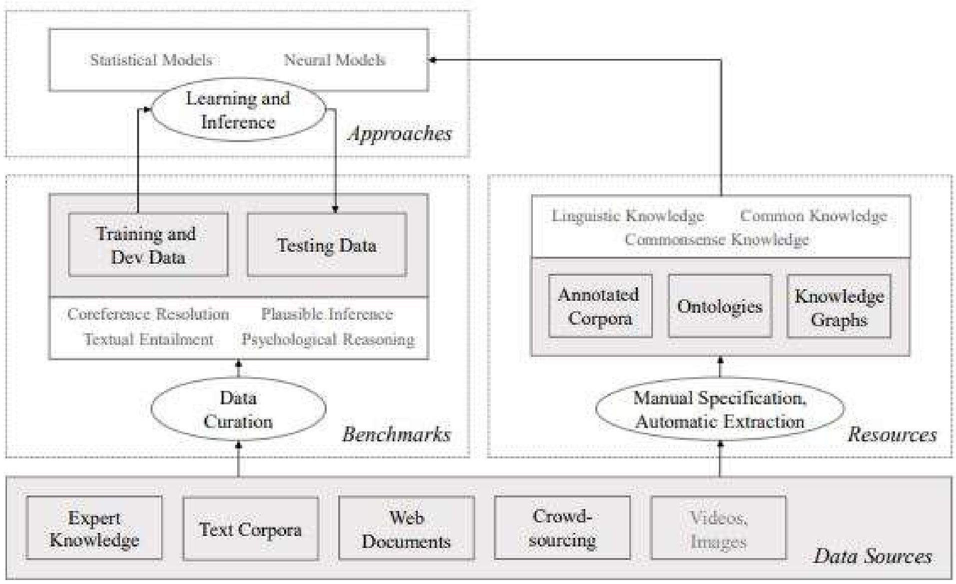
Davis & Marcus (2015) however, explained that notwithstanding these outstanding performances and advances in a multiplicity of NLP tasks, it's still imprecise whether these models are performing complex reasoning, or if they are simply learning complex surface correlation patterns (Marcus, 2018). Ye, Chen, Wang, & Ling (2020) have proposed a pre-training method for integrating commonsense knowledge into language representation models where they built a commonsense-related multi-choice question answering dataset to be used for pre-training a neural language representation model. Tandon, Varde, & Melo (2017) believe that mining knowledge and commonsense from huge amounts of data and applying it in intelligent systems, in many ways, seems to be the subsequent edge in computer science where they briefly presented an overview of the state of Commonsense Knowledge in Machine

Intelligence offers insights into Commonsense Knowledge acquisition, Commonsense Knowledge in natural language, applications of Commonsense Knowledge and conversation of related issues.

A very good survey on NLP was carried out by Gupta (2014) where he concluded that recent research in NLP shows more interest on learning algorithms which could be either semi-supervised or unsupervised in nature and available tasks of NLP are mostly: morphological separation, discourse analysis, natural language generation and understanding, machine translation, tagging of part of speech, recognition of named entities, optical characters recognitions, recognition of speech and analysis of sentiments etc.

Davis & Marcus (2015); Marcus (2018) have done an excellent job in providing a detailed account ranging from troubles in understanding and framing knowledge and commonsense reasoning for definite or general domains to difficulties in various forms of reasoning and their assimilation for the purposes of problem solving. Another interesting survey on commonsense knowledge reasoning for natural language understanding has been carried out by Storks, Gao, & Chai (2019) and the survey categorized commonsense knowledge and reasoning from the NLP community into three areas: benchmarks and tasks, knowledge resources, and learning and inference approaches as shown in the below figure.

Fig. 1.




Storks, Gao & Chai (2019)'s main research efforts in commonsense knowledge and reasoning from the NLP community in three areas: benchmarks and tasks, knowledge resources, and learning and inference approaches.

## 2. Overview of Existing Benchmarks

**Scientific and Practical Cyber Security Journal (SPCSJ) 6(2): 23-29 ISSN 2587- 4667 Scientific Cyber Security Association (SCSA)**

The NLP community has an extensive history of forming benchmarks to simplify algorithm development and evaluation for language processing tasks such as question answering, coreference resolution and named entity recognition (Storks, Gao, & Chai, 2019). Storks et. al. (2019) gave a review of broadly used benchmarks, presented by the subsequent groupings: textual entailment, question answering, plausible inference, multiple tasks, coreference resolution and psychological reasoning.

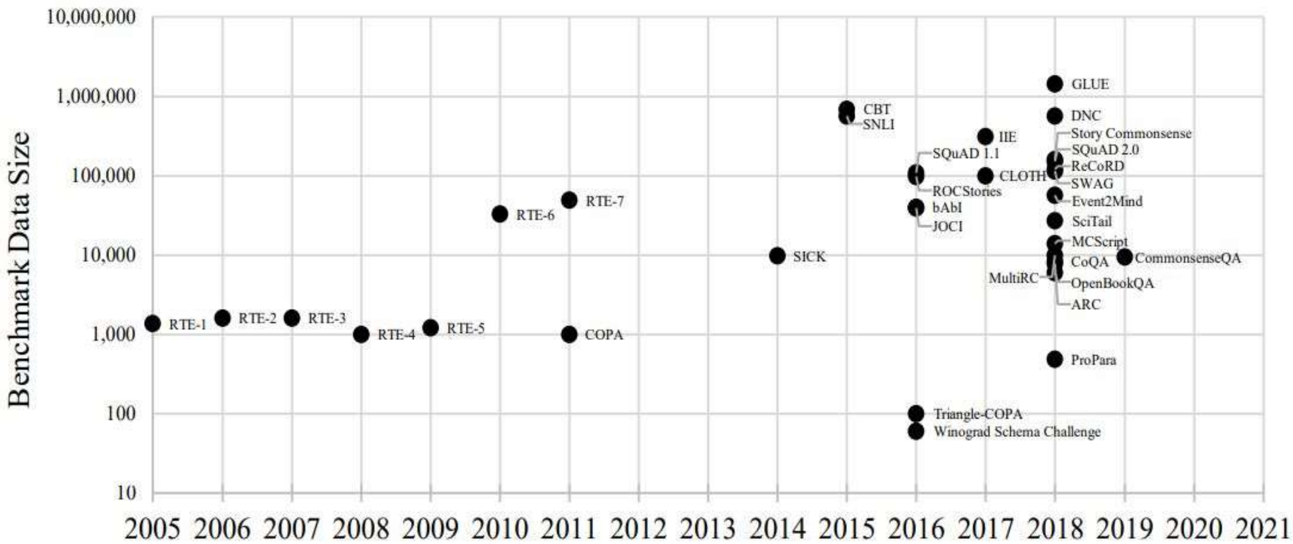



Fig. 2. Benchmark tasks between 2005 to 2021 geared towards commonsense reasoning for Natural Language Processing by (Storks, Gao, & Chai, 2019).

Ruder (2021) has defined a benchmark as it is used in NLP normally with numerous components: it entails of one or multiple datasets, one or multiple related metrics, and a means to aggregate performance where he stated that in order to continue making improvement, there is need to update and refine the metrics, to replace efficient simplified metrics with application-specific ones. The recent GEM benchmark, for instance, explicitly includes metrics as a component that should be improved over time, as shown in the figure below.

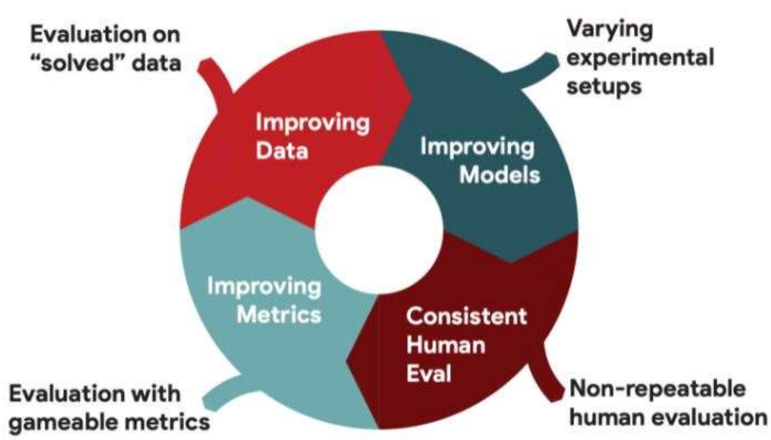


Fig. 3 Circle of challenges and opportunities of benchmark evaluation by (Gehrmann, et al., 2021).

A typical example of benchmark is the Rainbow Commonsense Reasoning benchmark. Rainbow is a worldwide commonsense reasoning benchmark spanning both social and physical common sense that brings together six existing commonsense reasoning tasks: aNLI, Cosmos QA, HellaSWAG, Physical IQa, Social IQa, and WinoGrande (Lourie, Bras, Bhagavatula, & Choi, 2021).

## 3. Knowledge Recourses

To understand human language, it is important to have linguistic knowledge resources that allow computers to identify syntactic and semantic structures from language and these structures in several cases need to be augmented with commonsense knowledge and common knowledge in order to reach a full understanding (Storks, Gao, & Chai, 2019). Chklovski (2003) has estimated that a typical human has accumulated numerous million diverse axioms of commonsense by adulthood. Baud, et al. (1996) have published a paper aiming at reviewing the problem of feeding Natural Language Processing (NLP) tools with convenient linguistic knowledge in the medical domain were he explained that a syntactic approach lacks the potential to solve a number of typical situations with ambiguities and is clearly insufficient for quality treatment of natural language.

Baud, et al. (1996) concluded that all the knowledge sources mentioned in his paper - together with others of course - are useful for NLP and when mining knowledge from various sources one is confronted with the problem of multiple or incompatible representation and one way to apparently solve this problem is to add another representation at the risk of augmenting the confusion for future users.

Some of the linguistic knowledge resources are:

I. Annotated linguistic corpora (Marcus, Santorini, & Marcinkiewicz, 1993)
II. Lexical resources. by (Miller, 1995)

Some of the common knowledge resources are:

i. YAGO by (Suchanek, Kasneci, & & Weikum, 2007)
ii. DBpedia by (Auer, et al., 2007)
iii. WikiTaxonomy by (Ponzetto & M, 2007)
iv. Freebase by (Bollacker, Evans, Paritosh, Sturge, & Taylor, 2008)

Some of the commonsense knowledge resources are:

i. Cyc by (Lenat & Guha, 1989)
ii. ConceptNet from (Liu & Singh, 2004)
iii. AnalogySpace (Speer, Havasi, & Lieberman, 2008)
iv. SenticNet by (Cambria, Speer, Havasi, & & Hussain, 2010)
v. ATOMIC by (Sap, et al., 2019)

There are several approaches to creating knowledge resources ranging from manual encoding to web documents text mining and crowdsourcing collection (Davis & Marcus, 2015) (Storks, Gao, & Chai, 2019).

### 3.1 Knowledge and Commonsense SOTA

There are several recent commonsense reasoning datasets that motivated researches in several aspects and domains which include: temporal, abductive, physical and social (Bhagavatula, et al., 2020). According Brown, et al. (2020) SOTA for most of them have achieved the close to human accuracy of over 80%. Conversely, their success is said to be due to larger pre-trained corpora as well as much more parameters, which would be challenging to be followed for most researchers.

## 4. Inference Approaches

Subsequently, Natural Language Inference is considered a benchmark task for testing the natural language understanding ability of the model by GLUE, Natural Language Inference has been well researched, and the language models have attained performance beyond humans on some Natural Language Inference datasets (Huang, He, & Liu, 2021). Additionally, by influencing transfer learning from large Natural Language Inference datasets, great performances have been achieved in numerous tasks, like in story ending prediction Li, Ding, & Liu (2019) and intent detection (Zhang, et al., 2020).

Huang, He, & Liu (2021) have proposed a framework that converts various commonsense reasoning tasks to a common task, Natural Language Inference and used a pre-trained language model, RoBERTa in solving it. By influencing transfer learning from large Natural Language Inference datasets, QNLI and MNLI, and adding vital knowledge from some knowledge bases like ATOMIC and ConceptNet, and their framework achieved state of the art unsupervised performance on the two commonsense reasoning tasks: CommonsenseQA and WinoWhy. Results from the experiments show that knowledge from QNLI and extracted from either ConceptNet or ATOMIC can complement one another to improve the model's performance on commonsense reasoning.

## 5. Conclusion and Recommendations

This survey explores the impact and importance of knowledge and commonsense reasoning in NLP. The survey provided a synopsis of the benchmarks, knowledge resources, state of the art and inference approach toward knowledge and commonsense reasoning for NLP. Devlin, Chang, Lee, & Toutanova (2019); Liu, et al. (2019), were the researchers who lately, made lots of advances in large pre-trained language models where they tried pushing machines nearer to humanlike understanding capabilities, making other researchers wonder if machines could directly model commonsense through symbolic integrations. Ye, Chen, Wang, & Ling (2020) have also proposed a pre-training method for integrating commonsense knowledge into language representation models where they built a commonsense-related multi-choice question answering dataset to be used for pre-training a neural language representation model. Storks, Gao, & Chai (2019) categorized commonsense knowledge and reasoning derived from the NLP community into three areas: benchmarks and tasks, knowledge resources, and learning and inference approaches as shown in the below figure. According Brown, et al. (2020) SOTA for most of them have achieved the close to human accuracy of over 80% but their success is said to be due to larger pre-trained corpora as well as much more parameters, which would be challenging to be followed for most researchers. Huang, He, & Liu (2021) have proposed a framework that converts various commonsense reasoning tasks to a common task, Natural Language Inference and used a pre-trained language model, RoBERTa in solving it.

There is need to revisit many implicitly accepted benchmarking practices such as depending on simplistic metrics such as BLEU and F1-score to keep up with improvements in modelling through taking motivation from real-world applications of language technology and considering the constraints and requirements that such settings pose for the models as recommended by (Ruder, 2021). Another emphasis should be put more rigorously in the assessment of models and rely on multiple metrics and statistical importance testing, contrary to present trends.

## REFERENCES

1. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2007). DBpedia: A Nucleus for a Web of Open Data. *The Semantic Web Challenge* (pp. 722-735). Busan, Korea: Springer Berlin Heidelberg.
2. Baud, R. H., Rassinoux, A. M., Lovis, C., Wagner, J., Griesser, V., Michel, P. A., & Scherrer, J. R. (1996). Knowledge sources for Natural Language Processing. *Proceedings of the AMIA Annual Fall Symposium* (pp. 70-74). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2233211/.
3. Bhagavatula, C., Bras, R. L., Malaviya, C., Sakaguchi, K., Holtzman, A., Rashkin, H., . . . Choi, Y. (2020). Abductive commonsense reasoning. *8th International Conference on Learning Representations, ICLR* (pp. 26-30). Addis Ababa: OpenReview.net.
4. Bollacker, K., Evans, C., Paritosh, P., Sturge, T., & Taylor, J. (2008). Freebase: A Collaboratively Created Graph Database for Structuring Human Knowledge. *The 2008 ACM SIGMOD International Conference on Management of Data* (pp. 1247–1250). NY, USA: SIGMOD .
5. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., . . . Child, R. (2020). Language models are few-shot learners. In Advances in Neural Information Processing Systems. *33: Annual Conference on Neural Information Processing Systems 2020.* Virtual: NeurIPS 2020.
6. Cambria, E., Speer, R., Havasi, C., & & Hussain, A. (2010). SenticNet: A Publicly Available Semantic Resource for Opinion Mining. *AAAI Fall Symposium on Commonsense Knowledge.* Menlo Park, CA, USA: AAAI Press.
7. Chklovski, T. (2003). Learner: A System for Acquiring Commonsense Knowledge by Analogy. *The 2nd International Conference on Knowledge Capture (K-CAP ’03), K-CAP ’03* (pp. 4-12). New York, NY, USA. : ACM.
8. Commonsenseknowledge. (n.d.). *http://commonsensereasoning.org*. Retrieved December 04, 2021, from www.commonsensereasoning.org: http://commonsensereasoning.org/
9. Davis, E., & Marcus, G. (2015). reasoning and commonsense knowledge in artificial Commonsense intelligence. *Commun. ACM*, 92–103.
10. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *In NAACL-HLT*.
11. Gehrmann, S., Adewumi, T., Aggarwal, K., Sasanka, P., Ammanamanchi, Anuoluwapo, A., . . . Emezue, C. (2021). *The GEM Benchmark: Natural Language Generation, its Evaluation and Metrics.* New York: Amelia R&D.
12. Gupta, V. (2014). A Survey of Natural Language Processing Techniques. *International Journal of Computer Science & Engineering Technology (IJCSET), 05*(01).
13. Huang, C., He, W., & Liu, Y. (2021). Improving Unsupervised Commonsense Reasoning Using Knowledge-Enabled Natural Language Inference. *Findings of the Association for Computational Linguistics*, 4875-4885.

14. Lenat, D. B., & Guha, R. V. (1989). *Building Large Knowledge-Based Systems; Representation and Inference in the Cyc Project.* Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.
15. Li, Z., Ding, X., & Liu, T. (2019). Story ending prediction by transferable BERT. *Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI* (pp. 1800-1806). Macao China: ijcai.org.
16. Liu, H., & Singh, P. (2004). ConceptNet — A Practical Commonsense Reasoning Tool-Kit. *BT Technology Journal*, 211-226.
17. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M. S., Chen, D., . . . Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach.
18. Lourie, N., Bras, R. L., Bhagavatula, C., & Choi, Y. (2021). *Rainbow: A Commonsense Reasoning Benchmark*. Retrieved 12 06, 2021, from www.allenai.org: https://allenai.org/data/rainbow